

**EVALUATING THE QUALITY OF ENGLISH FINAL EXAMINATION
OF THE SECOND SEMESTER TEST BY USING ITEMAN
SOFTWARE PROGRAM AT SECOND GRADE IN MADRASAH
TSANAWIYAH HASANUDDIN BANDAR LAMPUNG**

**A Thesis
Submitted as a Partial Fulfillment of
the Requirements for S1-Degree**

**By:
NURJANNAH
NPM.1711040115**

**Study Program : English Education
Advisor : Rohmatillah, M.Pd
Co-Advisor : Nurul Puspita, M.Pd**



**TARBIYAH AND TEACHER TRAINING FACULTY RADEN
INTAN STATE ISLAMIC UNIVERSITY
LAMPUNG
2020/2021**

ABSTRACT

EVALUATING THE QUALITY OF ENGLISH FINAL EXAMINATION OF THE SECOND SEMESTER TEST BY USING ITEMAN SOFTWARE PROGRAM AT SECOND GRADE IN MADRASAH TSANAWIYAH HASANUDDIN BANDAR LAMPUNG

**By:
NURJANNAH**

Teaching and learning is a process to interact the students toward their aims at level of learning experiences. To evaluate the quality of English final examination of the second semester test made by English teachers at Madrasah Tsanawiyah Hasanuddin Bandar Lampung using ITEMAN program software 4.4, based on the criteria of a good test.

The design of this research was descriptive-qualitative research which described the result of evaluation on an object based on standard criteria using ITEMAN Program Software. The instrument used in this research was students' test and students' answers from multiple choice test questions given by the teacher.

The finding of the analysis indicated the test items had low validity, low reliability, high discrimination power and average level of difficulty. It meant that the test items included the test that has not fully standardized of the good test for the students and the teacher had to revise the item test to make the test items become appropriate to use. Then the teachers should be able to make the appropriate test items for the students using a tool to measure the quality of item test. The teachers should be trained how to analyze the test items effectively and efficiently.

Keyword : Evaluating Test, Iteman Program Software, Final Examination

DECLARATION

I declared that this graduate thesis entitled “Evaluating the Quality of English Final Examination of the Second Semester Test by Using IteMan Software Program at Second Grade in Madrasah Tsanawiyah Hasanuddin Bandar Lampung” is entirely my work. I am accurately aware of the fact that I have quoted some statements and ideas of various sources and I duty acknowledged them in this thesis.

Bandar Lampung, May 21st 2021

The Researcher



NURJANNAH
NPM. 1711040115



KEMENTERIAN AGAMA
UNIVERSITAS ISLAM NEGERI RADEN INTAN LAMPUNG
FAKULTAS TARBIYAH DAN KEGURUAN

Alamat: Jl. Letkol H. Endro Suratmin Sukarame Bandar Lampung Telp. 0721-780887 fax.0721-780422

APPROVAL

Title : Evaluating the Quality of English Final Examination of the Second Semester Test by Using Iteman Software Program at Second Grade in Madrasah Tsanawiyah Hasanuddin Bandar Lampung.

Student's Name : Nurjannah
Students's Number : 1711040115
Study Program : Pendidikan Bahasa Inggris
Faculty : Tarbiyah and Teacher Training

APPROVED

Was Tested and defended in the examination session
at Tarbiyah and Teacher Training Faculty, State Islamic University
Raden Intan Lampung

Advisor

Rohmatillah, M.Pd
NIP. 198105082007102001

Co-Advisor

Nurul Puspita, M.Pd
NIP. 198907172015032004

The Chairperson of
English Education Study Program

Maisuri, M.Pd
NIP. 198005152003122004



**KEMENTERIAN AGAMA
UNIVERSITAS ISLAM NEGERI RADEN INTAN LAMPUNG
FAKULTAS TARBIYAH DAN KEGURUAN**

Alamat : Jl. Let. kol. H. Endro Suratmijn Sukarame Bandar Lampung Telp. (0721)703260

ADMISSION

A research proposal entitled: **Evaluating the Quality of English Final Examination of the Second Semester Test by Using Iiteman Software Program at Second Grade in Madrasah Tsanawiyah Hasanuddin Bandar Lampung**, was tested and defended in the examination session held on: Tuesday, August 31st 2021.

Board of Examiner:

Moderator : Meisuri, M.Pd

Secretary : Sri Suci Suryawati, M.Pd

Primary Examiner : Nunun Indrasari, M.Pd

Advisor : Rohmatillah, M.Pd

Co-Advisor : Nurul Puspita, M.Pd

**The Dean of
Tarbiyah and Teacher Training Faculty**



Prof. Dr. Hj. Nirva Diana, M.Pd

NIP. 196408281988032002

MOTTO

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

فَإِنَّ مَعَ الْعُسْرِ يُسْرًا ﴿٥﴾ إِنَّ مَعَ الْعُسْرِ يُسْرًا ﴿٦﴾ فَإِذَا فَرَغْتَ فَانصَبْ ﴿٧﴾ وَإِلَىٰ رَبِّكَ فَارْغَبْ ﴿٨﴾

In the name of Allah, the Most Gracious and the Most Merciful

“For indeed, with hardship [will be] ease. (5) Indeed, with hardship [will be] ease. (6) So when you have finished [your duties], then stand up [forworship]. (7) and to your Lord direct [your] longing. (8)”
(Q.S Al Insyira, 94: 5-8)¹



¹ Al Qur'an 3 Bahasa: Arab, Indonesia, Inggris, (Depok: Al Huda Kelompok Isnani, 2009) Al- Kolam: 1, pp

DEDICATION

All praises and gratitude to pray the presence of Allah SWT who has provided various kinds of pleasures, strengths, blessings that always felt all times. With all my heart and full of happiness, the author dedicates this thesis to people who are never tired of loving, accompanying, supporting, praying, and giving the meaning of the spirit of life not to give up easily, which in turn builds the author's personality so that self-awareness must continue learning Better yet, harder and harder every time, including:

1. My two beloved parents, Mr Jasman and Ms Maftuhah, who always give me sincere love and affection, who sincerely care for, educate and take care of me until now, and always provide prayer, encouragement, support, and motivation so that the author can complete education at the State Islamic University of Raden Intan Lampung, whose services the author could not reply to. May Allah always give His grace, health, ease of sustenance, blessings of age, and always be protected by Allah SWT.
2. For my dear brother, Bahrul Hidayat and my two younger sisters, Fatimah and Istiqomah who I really care about, who always give me encouragement and support.
3. All my extended family and siblings who have given their prayers and support.
4. Beloved Almamater of Raden Intan Lampung State Islamic University, a place to study and gain knowledge, hopefully it will become a better university in the future.

CURRICULUM VITAE

Nurjannah was born in Bandar Lampung April 21th 1999. She is the second daughter of four children from the couples Mr. Jasman and Mrs. Maftuhah. She has three siblings, one brother named Bahrul Hidayat and two sisters named Fatimah and Istiqomah.

She began his education at SD N 1 Kupang Raya Bandar Lampung in 2005, and graduated in 2011. Then she continued his education at the Junior High School level at Madrasah Tsanawiyah Hanuddin Bandar Lampung, graduated in 2014. In 2014 she continued the study at Senior High School in SMA Negeri 10 Bandar Lampung, which graduated in 2017. Then in 2017 she continued the level of Higher Education at State Islamic University of Raden Intan Lampung at the Tarbiyah and Teaching Faculty of English Education Study Program.

In 2020 she took part in Program Kuliah Kerja Nyata (KKN-DR) program in Penengah Raya Village, Kedaton District, Bandar Lampung. Furthermore, following Praktik Pengalaman Lapangan (PPL) activity at MTs Hasanuddin Bandar Lampung.



ACKNOWLEDGEMENT

Bismillahirrahmanirrahim

Alhamdulillahillobbil Alaamin, Gratitude is given for the presence of Allah SWT who has bestowed His grace and guidance in the form of knowledge, guidance, health, so that the author can complete the thesis entitled **"Evaluating the Quality of English Final Examination of the Second Semester Test by Using IteMan Software Program at Second Grade in Madrasah Tsanawiyah Hasanuddin Bandar Lampung "**. Hopefully prayers and greetings are still bestowed on the Prophet Muhammad SAW and also his family, friends, and people who are always on His path.

This thesis is a part and requirement for completing the undergraduate (S1) education study program at Tarbiyah and Teacher Training Faculty, English Education Study Program, State Islamic University of Raden Intan Lampung, in order to obtain a Bachelor of Education (S.Pd) degree.


The author is fully aware that this thesis is not successful without the guidance, assistance, support, and facilities provided. For that with all humility, and full of respect and thanks the authors say to:

1. Prof. Mrs. Dr. Hj. Nirva Diana, M.Pd, as Dean of the Faculty of Tarbiyah and Teacher Training, State Islamic University of Raden Intan Lampung
2. Ms Meisuri, M.Pd, as the head of the English Education Department of State Islamic University of Raden Intan Lampung
3. Ms Yulan Puspita Rini, M.A, as the Secretary of the Department of State Islamic University of Raden Intan Lampung
4. Ms Rohmatillah, M.Pd, as the First Advisor who has provided guidance and directed sincerely and patiently in completing this thesis.
5. Ms Nurul Puspita, M.Pd, as the Second Advisor who has provided guidance and direction sincerely and patiently in completing this thesis.
6. All Lecturers, Employees, and All Staff Employees in the Faculty of Tarbiyah and Teacher Training, State Islamic University of Raden Intan Lampung.

7. Head of Regional Libraries and Libraries of UIN Raden Intan Lampung who have lent books, references in completing the thesis.
8. Beloved Almamater State Islamic University of Raden Intan Lampung, where the author studied and gained knowledge, hopefully one day it will continue to be a better university.
9. My thesis partner, Irham Maghfury, Nathasya, Nurul Ulfani, Nur Fadillah, Nur Retno Ariany, Indri Yani, Firda Zakiyah, Ika Febrina Margarini, Azizah Pulungan, Laras Intan Nadia, and also the big family of PBI Class C 2017 who have provided a lot of support, please do 'a, as well as assistance so that the author can finish writing this thesis.
10. All parties who cannot be mentioned one by one who have contributed both morally and materially in the completion of the thesis.

The author would like to thank those who have helped this writing. Given their abilities, constructive criticism and suggestions from all readers are expected to improve further writing. Hopefully this thesis can be useful and become a good deed. *Aamiin Ya Rabbal 'Alamin*

Bandar Lampung, May 21st 2021


Nurjannah

NPM.1711040115

TABLE OF CONTENTS

ABSTRACT	ii
DECLARATION.....	iii
APPROVAL	iv
ADMISSION.....	v
MOTTO	vi
DEDICATION.....	vii
CURRICULUM VITAE	viii
ACKNOWLEDGEMENT	ix
TABLE OF CONTENTS.....	xi
LIST OF TABLES	xiii
LIST OF APPENDICES.....	xiv
CHAPTER I INTRODUCTION.....	xv
A. Title Affirmation	1
B. Background of the Research.....	2
C. Focus and Sub Focus of the Research	7
D. Identification of the Research.....	7
E. Formulation of the Problem.....	8
F. Objective of Research.....	8
G. Uses of the Research	9
H. Relevance Studies.....	10
I. Research Methodology.....	11
1. Research Design	11
2. Research Subject.....	12
3. Instrument and Data Collection	13
4. Research Procedure.....	14
5. Data Analysis.....	15
6. Trustworthiness of Data.....	17
J. Systematic of Discussion.....	18
CHAPTER II REVIEW OF RELATED LITERATURE	21
A. Frame of Theory	21
1. Language Test.....	21
2. Criteria of a Good Test Item	26
3. Iteman Software Program	29

4. English Language Skill Test	30
CHAPTER III DESCRIPTION OF RESEARCH OBJECT..	35
A. Research Object Description	35
B. Presentation of Research Fact and Data	35
CHAPTER IV RESULT AND DISCUSSION	39
A. Result of the Research	39
1. Iteman Software Program	39
2. Criteria of a Good Test.....	41
3. Revision Should be Made	58
B. Discussion.....	61
CHAPTER V CONCLUSION AND SUGGESTION.....	65
A. Conclusions	65
1. Item Software Program	65
2. Criteria of a Good Test Using Iteman Program Software 4.4	65
3. Revisions Should be Made by the Teacher	66
B. Suggestions	66
1. Suggestions for the Teacher	66
2. Suggestions for the students.....	67
APPENDICES.	

LIST OF TABLES

Table 1.1 Criteria of Test Item Quality	9
Table 1.2 Percentage of Evaluating the Quality of English Final Examination	20
Table 1.3 Calculation of the Percentage	21
Table 1.4 Example of Table Explanation Catagory	21
Table 1.5 Specification of Iteman Software Program	22
Table 1.6 Result of Reliability Analysis	24



LIST OF APPENDICES

Appendix 1 The Questions of Final Examination Grade 8 th	41
Appendix 2 Interview Guidelines the Teachers in Preliminary Research	47
Appendix 3 Transkrip of Interview Result the Teacher in Preliminary Research	48
Appendix 4 Application Research Letter from UIN Raden Intan.....	49
Appendix 5 Application Research Letter from School.....	50
Appendix 6 Preface Validation.....	51
Appendix 7 Validation Letter	52
Appendix 8 Syllabus Grade 8 Madrasah Tsanawiyah Hasanuddin	53
Appendix 9 The Students' Answer of Final Examination Grade 8 th	71
Appendix 10 The Key Answer of Final Examination Grade 8 th	72
Appendix 11 The Summary of Statistic of Final Examination Grade 8 th	73
Appendix 12 Item Statistic of the Final Examination Grade 8 th ..	74
Appendix 13 The Result Percentage of Analysis	87
Appendix 14 The Photo of Item Program Software 4.4	88
Appendix 15 The Document of Iteman Program Software 4.4 ..	89
Appendix 16 The Photo of Interview with English Teachers Grade 8 th	91
Appendix 17 Control Card.....	92
Appendix 18 Result of Turnitin.....	93

CHAPTER I INTRODUCTION

A. Title Affirmation

In understanding the title of this research and avoid misunderstanding, the researcher needed to explain this research.

The evaluating referred to this research was evaluating the English final examination test at the second semester for grade eight, based on the criteria of a good test identified of validity, reliability, discrimination power, and level of difficulty. The questions evaluated are questions of the multiple choice test are 50 questions

The quality referred to this research was the quality of the questions tested to the students during the final examination test based on the criteria of a good test, to find out whether the questions are feasible or not based on the eligibility criteria for the questions.

The final examination referred to this research was the final examination for grade eight in the second semester for class advancement to a higher grade, namely towards grade nine in academic year 2020.

Evaluating quality of final examination in this research was an evaluation of the final examination in formed multiple choice tests for 50 questions made by the teacher to indicate the questions are suitable or not for the students while in pandemic era.

Iteman software program that used in this research was ITEMAN 4.4, which was a program used to analyze items and analyze test results. The focus of the researcher was to measure the validity, reliability, discrimination power and level of difficulty in each item for the second semester of the final examination using iteman software program 4.4.

Madrasah Tsanawiyah Hasanuddin referred to this research was a junior high school located in Bandar Lampung, precisely in Kupang Teba, North Teluk Betung , Bandar Lampung.

The researcher means by the title of the research about **Evaluating the Quality of English Final Examination of the Second Semester Test by Using Iteman Software Program at Second Grade in Madrasah Tsanawiyah Hasanuddin Bandar Lampung** was to conduct analysis items for each questions of the final examination using iteman program software.

B. Background of the Research

Teaching and learning is a process to intract the students toward their aims and combine the knowledge, behaviors, and skills that add their level of learning experiences. Learning process is a process that people pass through the new knowledge, capability and skills then ultimately influence their attitude and actions. The learning will happen when people can demonstrate that they know something that they did not know something before and when they can do something they could not do before.² To find out the result of their learning, test is the choice for taking their experiences in learning. It is one of the ways to measure students' ability in learning the subject. Test is designed to measure students' ability in a language regardless of the training that they may have had in that language³. The test is used as a tool to measure the students' understanding of the learning that the teacher has given to the students while studying.

Test is as one of the procedure the teacher needs to analyze students to see what they need and want. According to Hughes "Test is used to obtain the information that we hope to

² Honey , P and Mumford, A, *The Manual of Learning Style*, 3rd edn (Maidenhead: Honey Publication, 1996)

³ Hughes Arthur, *Testing for Language Teachers* (New york: Cambridge University Press,1989), p.9

obtain will of course vary from situation to situation”⁴ Test is implied by the questions to apply the teacher’s questions is used to make a questions to avoid the uncontrol test based on the criteria, it is why they cannot be separated.

There are many types of tests that a student can take to gauge his or her fitness in a single study, such as having a daily test, oral test, monthly test, mid test or final test. Teacher can diagnose the students’ strengths and weaknesses in oral or written communication.⁵ That is implied to see how far the students understand the learning of study. A test is a method to measure the persons’ ability of knowledge, or performance by given domain.⁶ Teacher can give test as the implementation of students’ result of what they listened and comprehended from the teacher while they were studying.

To know the quality of items in a test based on the result students’ test that have given, teacher can do evaluation. Evaluation is the way to evaluate the questions as the teachers’ references and review. After getting the evaluation, teachers are able to comprehend all materials start from the students, level of students, the difficulties of students and also the questions. Bachman defines evaluation “as the systematic gathering of information to make decisions”⁷. Evaluation can be used to measure the students’ skill on any subject accurately.

There are a lot of ways to measure the students’ abilities, one of the tests is multiple choice test. Some teachers choose multiple choices as a test, because it can become efficiently and effectively way. Multiple choice tests supposed as the most standardized test in testing the students’ ability and knowledge based on the choices of the answers. Multiple choice test is

⁴Ibid.,

⁵ Madsen S. Harold, *Techniques in Testing* (New York: Oxford University Press, 1983), P.11

⁶ Brown H. Douglas, *Language Assessment: Principle and Classroom Practices* (America: Longman, 2004), p.3.

⁷ Bachman, L. F, *Fundamental Considerations in Language Testing* (Oxford University Press, 1990), p.21.

suitable as a strategy in assessment.⁸ Teacher indicates that multiple-choice tests can be useful and it is the right test for measuring the knowledge of the students so that they can analyze the material.

Multiple choice is the test that is formed from the questions and the close ended answer. The examinees just need to choose the best option from the available choices. Multiple choice can be as *productive measure* like speaking test, require active or creative answer, *receptive measure* like multiple reading test, tent to rely on recognition, with the students simply choosing the letter of the answer.⁹ Skill in English can be formed to the multiple choices with the provisions in reading skill.

In learning English, reading skills determine the important contributions. From reading skill learners can receive inputs for the development of listening, speaking, and writing skills through reading.¹⁰ They can help the learners to gain the knowledge of speaking expression, writing technique, or comprehension in listening through printed texts deeply. Learners also can get more grammar and vocabulary that introduced in reading.

Test should measure four aspects of the criteria of a good test such as validity, reliability, level of difficulty and discrimination power. Many teachers still use manual way to analyze the quality of the test items, it will spend a lot of time. So as many teachers use the media as a tool to help them evaluate the results of student questions and facilitate them in a short time and efficient. Properly the teachers can use ITEMAN software program to analyze the quality of the final examination test items that they have made.

⁸ O'Dweyer Aidar, "A Teaching Practices Review of the Use of Multiple Choice Questions for Formative and Summative Assessment of Student Work on Advanced Undergraduate and Postgraduate Modules in Engineering". *Jurnal AISHE*, Vol 4 No 1(Spring 2012), p.10

⁹ Madsen S. Harold, *Techniques in Testing* (New York: Oxford University Press, 1983), P.12

¹⁰ Fiharsono A, *An Analysis of the Use of Multiple Choice Items of Reading Test in UAN for Students of SMP* (Yogyakarta: Sanata Dharma University, 2005)

ITEMAN software can be used by the teacher to analyze test items and decide which test item is good to use and not to use. Based on the validity, reliability, discriminating power, and level of difficulty. It is why this program software is important for the teachers because this software program can be administering the test items in simple way.

Based on the explanation above, the problem concerning the analysis of test items needed to be investigated because the test items will be used for testing the students in the exam, the teachers have to analyze first. It is the reason why the researcher interested in analyzing the second Semester test items made by the teacher using ITEMAN program software 4.4, which would be used for Madrasah Tsanawiyah Hasanuddin Bandar Lampung. The aim of the item analyses was to establish the test items with the result of ITEMAN program software and the revisions for the items.

From the background of relation of this research, there are some studies that related to the quality of English examination test in each item generally. There are three researchers that conducted the research about the quality of English test in each items. They are Nofiyanti in 2011, Fatma Dwi Rusmiana in 2015, and Nurul Amanah in 2017.

Nofiyanti investigated the analyzing the content validity of English summative test for the first year students of junior high school at SMP N 87 Jakarta, and she focused on the validity test of the content validity. She clarified that the validity of test referred to the representativeness of the questions included in containing of the instrument to give each its emphasis. She said the content validity concerned with it whether or not the content of the test is representative and comprehensive sufficiently for the test to be valid the measurement of what supposed to measure.¹¹

¹¹ Nofiyanti, "An Analysis on the Content Validity of the Summative Test for the First Year Students of Junior High School". (The Research of the s1 Degree in English Language Education UIN Syarif Hidayatullah Jakarta, Jakarta, 2011), Pg. 7.

Fatma Dwi Rusmiana investigated that the analysis of the quality about questions of the 1st semester final test of the accounting theory vocational at class XII at SMK YPKK 1 Sleman Academic Year of 2014/2015 is good based on the result using the ITEMAN. She focuses on the result the quality of test items by ITEMAN.¹²

Nurul Amanah investigated that the analyzing the quality of final semester test by using ITEMAN at the second grade of SMP negeri 2 Sumberjo Tanggamus in 2016/2017. The focus of the research is the analyzing on the side and investigating the population which has different knowledgeable students and multiple choice items. She focuses on the evaluation of the quality that is used the ITEMAN and does not use it based on the result.¹³

From the related studies above the studies mentioned test as the measure of the progress in study for the students and ITEMAN as the tool to analyze the test items for junior high school or senior high school so that the program can be the optional for the teacher to evaluate the process teaching and learning use the result.

Some differences between the previous research and this research were the questions that used to analyze was the second grade of final examination in academic year 2020, the researcher concerned how to know the quality of English final examination of the second semester test by using ITEMAN software program software 4.4 at second grade in Madrasah Tsanawiyah Hasanuddin Bandar Lampung in pandemic condition. This research was to evaluate and analyze the test items of multiple choices using ITEMAN program software, and gave the revisions

¹² Fatma Dwi Rusmiana, "The Test Analysis of 1st Semester Final Test of the Accounting Theory for Vocational Education: Case Study of SMK YPKK 1 Sleman Academic Year of 2014/2015" . (The Research of Degree Bachelor of Education in Faculty of Economics Yogyakarta State University, Yogyakarta, 2015), Pg. 58.

¹³ Nurul Amanah, "The Analyzing the Quality of Final Semester Test by Using ITEMAN at the Second Grade Of SMP Negeri 2 Sumberjo Tanggamus in 2016/2017". (The Research of s1 Degree the Training and Education Faculty University of Lampung, Lampung, 2017), Pg.11.

for the questions based on the result of the ITEMAN program software.

C. Focus and Sub Focus of the Research

Based on the background of the problems above, this research focus was on the quality of test items for second Semester Examination of the Second semester at the second grade or eight in Madrasah Tsanawiyah Hasanuddin Bandar Lampung who go to the next level using Iteman software program. The questions that evaluated are multiple choice that usual students get when they are doing examination. Some English teachers did not understand and know how to analyze the test items that they had made before they used for the students during the examination test. It was why the students were rarely got good scores in the examination.

Moreover, the quality of test items were questionable because they did not analyze before. It should be the test items for the students' test had to be analyzed to know the test that they belonged based on the criteria to a good test or not. Therefore, this research would focus on the analyzing the kinds of the multiple choice test items only, however this program software could not be used for essay items.

The sub focus of this research was iteman program software 4.4 to know the validity, reliability, level of difficulty and discrimination power of the item for each question using the program.

D. Identification of the Research

According to the background of the research above, the following problems can be found:

1. The English Teacher of Madrasah Tsanawiyah Hasanuddin Bandar Lampung did not know what was iteman program software and the usage of iteman software.

2. The English teachers of Madrasah Tsanawiyah Hasanuddin Bandar Lampung who were teaching grade eight at the second semester were unfamiliar about the program software and rarely analyzed the test to determine of the quality of test items.
3. The English final examination of the second semester test items in Madrasah Tsanawiyah Hasanuddin Bandar Lampung were made by the teacher were not identified of the validity, reliability, discrimination power, and level of difficulty.

E. Formulation of the Problem

Based on the identification, the researcher formulates the formulation of the problem as follows:

1. How did iteman software program run and what else that could analyze using iteman analyze?
2. How was the quality of the English final examination of the second semester test made by English teachers based on criteria of a good test (validity, reliability, discrimination power, and level of difficulty)?
3. How revision should be made by the teacher on English final examination of the second semester test items in Madrasah Tsanawiyah Hasanuddin Bandar Lampung?

F. Objective of Research

In line with the formulation of the problem above, the objectives of this research are:

1. To tell the readers what was iteman program software and what else that could analyze using iteman.
2. To evaluate the quality of the English final examination of the second semester test made by English teachers at Madrasah Tsanawiyah Hasanuddin Bandar Lampung use ITEMAN

program software 4.4, the criteria of a good test (validity, reliability, discrimination power, and level of difficulty).

3. To know what revision should be made by the teacher on English final examination of the second semester test on second semester items in Madrasah Tsanawiyah Hasanuddin Bandar Lampung made by the teacher.

G. Uses of the Research

1. Theoretically

The results of this theory were expected to complete the previous theories of the quality of assessment.

2. Practically

This research might be used to assess the quality of multiple choice tests by using ITEMAN Program Software. The findings of this research were hopefully could be useful for many objects.

- a. For the teachers, this research might be used to help the teachers assess the quality of multiple choice tests by using ITEMAN Program Software.
- b. For the students, as the tool to measure the students' ability in learning and as the students' evaluation to change the way of study.
- c. For the school, the research hoped this ITEMAN could be used as the tool to evaluate the quality of questions and to evaluate the learning and teaching process based on the result of the program.
- d. For the other research, the results of this theory were expected to complete the previous theories of the quality of assessment.

H. Relevance Studies

The relation to this research, there were several studies related to the quality of English test items. There were three researchers who conducted the research, as follows:

Nofiyanti investigated the analyzing the content validity of English summative test for the first year students of junior high school at SMP N 87 Jakarta, and she focused on the validity test of the content validity. She clarified that the validity of test referred to the representativeness of the questions included in containing of the instrument to give each its emphasis. She said the content validity concerned with it whether or not the content of the test is representative and comprehensive sufficiently for the test to be valid the measurement of what supposed to measure.¹⁴

Fatma Dwi Rusmiana investigated that the analysis of the quality about questions of the 1st semester final test of the accounting theory vocational at class XII at SMK YPKK 1 Sleman Academic Year of 2014/2015 is good based on the result using the ITEMAN. She focuses on the result the quality of test items by ITEMAN.¹⁵

Nurul Amanah investigated that the analyzing the quality of final semester test by using ITEMAN at the second grade of SMP negeri 2 Sumberjo Tanggamus in 2016/2017. The focus of the research is the analyzing on the side and investigating the population which has different knowledgeable students and multiple choice items. She focuses on the evaluation of the quality that is used the ITEMAN and does not use it based on the result.¹⁶

¹⁴ Nofiyanti, "An Analysis on the Content Validity of the Summative Test for the First Year Students of Junior High School". (The Research of the s1 Degree in English Language Education UIN Syarif Hidayatullah Jakarta, Jakarta, 2011), Pg. 7.

¹⁵ Fatma Dwi Rusmiana, "The Test Analysis of 1st Semester Final Test of the Accounting Theory for Vocational Education: Case Study of SMK YPKK 1 Sleman Academic Year of 2014/2015". (The Research of Degree Bachelor of Education in Faculty of Economics Yogyakarta State University, Yogyakarta, 2015), Pg. 58.

¹⁶ Nurul Amanah, "The Analyzing the Quality of Final Semester Test by Using ITEMAN at the Second Grade Of SMP Negeri 2 Sumberjo Tanggamus in 2016/2017". (The Research of s1 Degree the Training and Education Faculty University of Lampung, Lampung, 2017), Pg.11.

From the related studies above the studies mentioned test as the measure of the progress in study for the students and ITEMAN as the tool to analyze the test items for junior high school or senior high school so that the program can be the optional for the teacher to evaluate the process teaching and learning use the result.

Some differences between the previous research and this research were the questions that used to analyze was the second grade of final examination in academic year 2020, the researcher concerned how to know the quality of English final examination of the second semester test by using ITEMAN software program software 4.4 at second grade in Madrasah Tsanawiyah Hasanuddin Bandar Lampung in pandemic condition. This research was to evaluate and analyze the test items of multiple choices using ITEMAN program software, and gave the revisions for the questions based on the result of the ITEMAN program software.

I. Research Methodology

1. Research Design

The design of this research was descriptive qualitative research which described the result of the evaluation on the object based on standard criteria using ITEMAN Program Software 4.4. Evaluation research was a the research that provide the information for the decision maker that related to the power or strength of a program software 4.4.¹⁷ Descriptive research studies were designed to obtain the nature of the situation as it exists at the time of the study.

This study used data that was already available. The purpose was to describe what exist.¹⁸ The research was intended to evaluate the questions using the ITEMAN

¹⁷ Arifin, Zaenal, Metodologi Penelitian Pendidikan Filosofi, Teori dan Aplikasinya, (Surabaya: Lentera Cendekia, 2010), P.15

¹⁸ Donald Ary, Introduction to Research in Education 4th (New York:Harcourt Brace Collage Publisers, 1990), P. 415

software program 4.4 and proposed several revisions based on the results obtained using the iteman analysis for the final semester exam in second grade at Madrasah Tsanawiyah Hasanuddin Bandar Lampung, located in Teluk Betung Utara, Bandar Lampung.

The Purpose of this research was to describe how was the usage of ITEMAN Program Software to analyze the final semester test of students in Madrasah Tsanawiyah Hasanuddin. The objects of this research conducted of the test items and the student's answer in soft file. Both of them were analyzed using the standard of criteria. It was the validity, reliability, level of difficulty, and discriminating power.

2. Research Subject

This research was conducted at the second grade of the final semester or second semester in Madrasah Tsanawiyah Hasanuddin Bandar Lampung, which was located in Teluk Betung Utara. This research was conducted over a period of a week. It was administered in English lesson which was tested after the students finished their English final semester test items.

The object of this research was the final semester test of English made by the teacher for the second grade at Madrasah Tsanawiyah Hasanuddin Bandar Lampung. The test questions given were in the type of multiple choice test which had multiple choices, namely choices with 4 choices, namely a, b, c, and d and had 1 choice as the answer key.

The number of test questions used in this research were 50 items. The test questions made by the teacher were tested to get data on student answers. The tests evaluated in this study were the answers of students who answered questions using an online test, totaling 20 students.

3. Instrument and Data Collection

a. Instrument

The instrument used in this research was student test as a data, namely teacher-made English final examination test items as documents to be analyzed, the number of questions in grade 2 at Madrasah Tsanawiyah Hasanuddin Bandar Lampung which amounted to 50 items and using student answers from the answer of multiple choice test questions given by the teacher.

The number of students who took the online test was 20 students, so that in this research it was found to have 20 student answers and 50 test items obtained from documents that were used as the basis for analyzing student examination questions using the software program iteman 4.4 as a tool to analyze questions based on student answers.

The Documents in this research were obtained by approaching the principal, curriculum teacher and English teachers. During the approach, the researcher would first ask the principal for permission to conduct research in the school, and then asked for permission to access the documents through the English teacher to request the documents needed for research, and the principal allowed it.

b. Data Collecting Technique

Final semester test or second semester test items were collected as the data for the research. The test items were analyzed to determine whether there would be several revisions for test items based on the result of the analysis while using ITEMAN Program Software 4.4 .

The data was collected by administering the teachers who made the questions for English final examination test items to the second grade of the final

semester in Madrasah Tsanawiyah Hasanuddin Bandar Lampung.

4. Research Procedure

To check the quality of the final examination test, there were some procedures to get the question that made by the teacher and the students' answers as the instruments in the final semester test. Each item had four options, they are A, B, C, and D. Then, the researcher analyzed the test using the iteman program software 4.4.

To make the research run well, there are several procedures as follows:

1. Determining the problems

The problems of the research were formulated as the foundation of this research.

2. Determining and selecting the population

The population that used of this research was the final semester test items at the second grade in Madrasah Tsanawiyah Hasanuddin Bandar Lampung.

3. Determining the class

The researcher used random sampling¹⁹, so the researcher took one class in this research, it was VIII A that used online test in Madrasah Tsanawiyah Hasanuddin Bandar Lampung. There were 20 students that used online tests.

4. Determining the test

The test was from the final test or second semester test. There were fifty multiple choice test items.

5. Asking the students' answers

After determining the test, the researcher asked the English teacher at grade eight to give the students' answer as the data to evaluate.

¹⁹ Arikunto, s, 2006, *Prosedur Penelitian Suatu Pendekatan Praktik Edisi Revisi VI*, Jakarta: Rineka Cipta. p.124

6. Analyzing the data quantitatively

This research touched the final examination test by counting the result of Iteman software program 4.4.

7. Analyzing the data descriptively

Final examination test in this research was identified by descriptive approach to find out the reliability, level of difficulty, discriminating power, and level of difficulty using iteman program software.

To analyze the data using iteman program, the following are the steps to enter the data using Iteman Program Software:

1. Open iteman program by clicking *start* to the futur
2. Select program/ *click iteman*
3. Type the name of the data file (input) as you want, then *Enter* the name of the input file. For example, D:\Finaltest.txt, then *Enter* again
4. Enter the name of the output file. For example, D:\Finaltest.output, then click *Enter*
5. A question will appear on the screen. The written is *Do you want the scores written to a file? (Y/N)*, then you can type *Y* and click *Enter*
6. Enter the name of your score file as you want: For example, D:\Finaltest.scr, then click *Enter*.
7. *Finish*.²⁰

5. Data Analysis

The data was collected by the researcher using administering the test of the teacher made for final examination test. The test that was administered by the researcher gether with the answer of the students'. The data

²⁰ Suparman, Ujang, Analyzing The Quality of English Test Items of Daily, Mid Semester and Final School Examinations in Bandar Lampung, A paper presented in International Conference on Educational Evaluation and Research (ICERE), May 29 and 30, 2016. P.4

analysis of this research was focused on the evaluating of the test items to find out the result of the the test items whether it was good or not, it could be seen at the points of: validity, reliability, discriminating power, and level of difficulty.

Whereas, the data of the analysis also was intended to determine the interpretation of the items that could be used well, should be totally revised, or partially revised, or dropped totally. To interpret the result of analysis using the iteman software program for the test items, there were several criterias of the quality of test.²¹

Table 1.1 Criteria of Test Item Quality

Level of Difficulty – p	
0.000 – 0.099	Very Difficult/need total revising
0.100 – 0.299	Difficult/needs revising
0.300 – 0.700	Average/good
0.701 - 0.900	Easy/needs revising
0.901 – 1.000	Very easy/needs dropping or totally revising
Point Biserial (Discriminating Power – D)	
$D \leq 0.199$	Very low/ needs dropping or totally revising
0.200 – 0.299	Low/needs revising
0.300 – 0.399	Quite average/without revision
$D \geq 0.400$	High/very good
Prop Endorsing (Proportion of the Answer)	
0.000 – 0.010	Low/needs revising

²¹ Suparman, U. (2011). The Implementation of Iteman to Improve the Quality of English Test Items as a Foreign Language (An Assesment Analysis). *Aksara Jurnal Bahasa*, Vol. XII. No. 1., Hal. 1-96. ISSN 1411-2501.

0.011 – 0.050	Sufficient/good enough
0.051 – 1.000	Very Good
Alpha (Test Item Reliability)	
0.000 – 0.400	Low /not sufficient
0.401 – 0.700	Average/sufficient
0.071 – 1.000	High/good

Criteria to classify the quality of test items²²

6. Trustworthiness of Data

The researcher needed to establish the validity of the score from qualitative measured and discussed the qualitative findings. The study referred to the degree of confidence in data, interpretation, and methods used to ensure the quality of a study²³. The trustworthiness must be done by the researchers to certain the data of the research.

In a descriptive-evaluative research used data triangulation to keep the result of the data accurate at the conclusion then to increase the validity and reliability of the result.²⁴ The researcher collected the data to assure the validity through the convergence of the information while the researcher uses different sources. To valid the data, credibility which members check were employed.

To describe the result of an evaluation of the object based on standard criteria might involve the number of data that become the result of the research using ITEMAN Program Software, it meant that the data collecting to convey

²² Ibid.10

²³ Colnelly, Lynne M, *Trustworthiness in Qualitative Research*, Medsurd Nursing, Vol.25 No. 6 (December 2016), P. 435

²⁴ Denzin, N, 1970, *Sociological Methods: A Source Book*, New Burnswick, Transaction Publisers, p.14

on the truth²⁵. The data identified the result of quality semester test using the software that had the keyword of level to classify the test items. So the method research appropriated to be used in this research.

The researcher used the data of students' answer as the measurement to know the quality of final examinations using ITEMAN program software, then the researcher analyzed the data based on the validity, reliability, discrimination power and level of difficulty using the result evaluation from the software. The last, researcher asked to the lecture selectively to check the accuracy of the researchers' conclusion.

J. Systematic of Discussion

The researcher discussed the research in the structure as below:

In Chapter I, the researcher needed to tell the general description of the research and the purpose of the research. It was why the researcher presented the introduction, which consisted of title affirmation, background of the problem, focus and sub focus of the research, identification of the problem, formulate the problem, objective of the research, uses of the research, relevance studies, research methodology and systematic discussion.

In Chapter II, the researcher needed to tell the related literature of the research and the theories that researcher used to do research as base of the research. It was why the researcher presented the theories of language test in English, quality of test items, iteman software program, criteria of a good English language skill test which consist of explanation in English test.

In Chapter III, the researcher needed to tell the description of the research to clarify the research. It was why the researcher presented the general description of evaluating the quality of English final examination of the second semester test by using iteman software program at second grade in Madrasah

²⁵ Polit, Desine F, *Essentials of Nursing Research*, (gold Coast: Wolters Kluwer Health, 2020), P.106

Tsanawiyah Hasanuddin Bandar Lampung and presentation of research facts and data.

In Chapter IV, the researcher needed to tell the result of findings research. It was why the research presented the result of evaluating the quality of English final examination of the second semester test by using iteman software program at second grade in Madrasah Tsanawiyah Hasanuddin Bandar Lampung which consists of analysis of research data, and research findings.

In Chapter V, the research needed to tell the conclusion of the research and the suggestions. It was why the researcher presented the conclusion and suggestion of the research.



CHAPTER II

REVIEW OF RELATED LITERATURE

A. Frame of Theory

1. Language Test

a. Definition of Test

Test is used to know the capability of person in comprehending something and to know the result of learning. Test is used to obtain the information that is hoped to vary from situation to situation²⁶. Test is the way that can establish the students to master the language, to know it when they study for exam and again when exams are returned and discussed.²⁷ Test is used to measure the ability of students and for gauge the students' knowledge of the lesson.

Test is used to measure the ability of students so it has purpose to make students detect their mistakes in doing test and know the reason of the problems. The aim of the test is to give the information to aid in making intelligent decision about the possible courses of action that did.²⁸ Test is as the informant about the overall materials of subject that they has learnt,

Test can be as the tool to illustrate the ability of the students, it includes strength and weakness of students' comprehension in understanding the subject that has already been studied while studying and as the way to know the result of students' learning while studying.

²⁶ Hughes Arthur, *Testing for Language Teachers* (New York: Cambridge University Press, 1989), p.9

²⁷ Madsen, Harold S, *Techniques in Testing* (New York: Oxford University Press, 1983), p.4

²⁸ Fulcher, Glenn, *Practical Language Testing* (Great Britain: Hachette UK Company, 2010), p.1

Test became the standar to measure the students' ability in comprehending the knowledge of the subject that has been learnt during the process of the learning, therefore the test is called as a tool that have to be in the learning.

b. Kinds of Test

According to the information that sought, the categorize that proved to determine about the existing test that is suitable for the particular purpose and the appropriate test that is necessary. There are severals types of test that discussed, they are:²⁹

1) Proficiency Tests

Proficiency tests are designed to measure the student's ability in getting the language regardless of the training that they might have for their language. In case proficiency test means getting the sufficient command about the language for the particular purpose. The example of proficiency test is test that designed to find who can get successfully in the function as a United Nation Translator or test that used to discover students who has good English to follow the study in Britain University.

2) Achievement Tests

Achievements test are related in the language course, the purpose is to establish how successful ownself, the groups of students, or the course that achieves the objectives. There are two types of the achievement tests, they are: *final achievement* tests that administrated at the final of the course of study that the students taking and *progress achievement tests* that intended to measure the progress that

²⁹ Hughes Arthur, *Testing for Language Teachers* (New York: Cambridge University Press,1989), p.9

students making. The tests of progress relate to the objective of the course, the way to measure the progress repeatedly to administer the final achievement test.

3) Diagnostic Tests

Diagnostic tests are used to identify students' strengths and weaknesses. they intended primary to certain the necessary of further teaching. At the level of the language skills this is reasonably straight forward. It can tell the students' particularly weak in speaking or writing in order to create on the profiles of the students' ability such categorize as grammatical accuracy or even the linguistic appropriacy. They are used to find student's difficulty in learning and to investigate what kind of difficulty that they face.

4) Placement Tests

Placement tests are provided about information that will place the students at their grade at the program class to appropriate their abilities. The usage is to assign the students level of class that will decide by the institution they taking. It might be designed within an institution is given.

The test that the researcher analyzed was the achievement test, because it could be seen that the researcher researched final examination test or in Indonesia usually called by UAS test. It is held to measure the capacity or the ability of the students in understanding the materials of the learning for that grade. Final examination or UAS test obtains about the materials that relevant in the learning and the syllables. The result, the test that used for this kind of test is achievement test.

c. Form of Test

From the form of the test, there are two kinds based on the response they are subjective test and objective test.³⁰

1) Subjective Test

Subjective test is the test that used to know the ability of the students by their opinion or the judgment. Kinds of tests that includes in subjective test is essay questions and short answer questions.

a) Essay Test

An essay test is the assessment test that has the technique requires the students' respond to the question by developing themselves, organizing the class of word or grammatical that used, and the written of the original composition. It is used to measure the students' abilities to construct the logical, the cohesive, and the the persuasive in written on pieces

b) Short Answer

Short-answer tests are constructed to the object items, the answers are clearly required on it. The answer has to be supplied on it or the options have clearly provide. It measures the items of the perfect reliability in the true objective. The answers provide the unquestionable whether right or wrong.

2) Objective Test

Objective test is a test that use to know the students' ability or their performance that measured by using the specific answers or the possible answers that provide the wrong and the right answers in the items.

³⁰ Hughes Arthur, *Testing for Language Teachers*, (New York: Cambridge University Press, 1989). P.22

There are several types of the objective test: multiple choice tests, matching and true false test.

a) Multiple Choice Tests

Multiple choice test consists of the questions with the answers about incomplete definition. Multiple choice consists of the problem and the lists of the suggested solutions.³¹ To answer the questions of multiple choice students have to choose one of the best possible answers.

Multiple choice contains the questions that called as *stem* with the possible answers that called *choices*, it also includes the right answers and some wrong answers that called *distractors*. The students have to answer by circle the one that they think right on the readable answer sheet.

The alternatives answers in the multiple choice consist of the complete sentences, sentence fragments, or the single words. The items assume absolutely correct.³² The advantage of multiple-choice is that the teachers score perfectly reliable of factual information.³³ The teacher score it easily and the students can guess he possible answer easily too.

b) True or false Test

True or false tests consist of the statements. In this type students can answer the questions the available on the instructions

³¹ Gronlund, N. E, *Measurement and evaluation in teaching*, (New York: Mac. Milan Co, 1967).

³² Osterlind, S, *Constructing test items and Multiple choice, constructed response, performance and other formats*, (New York: Kluwer Academic Publishers, 1998), p20.

³³ Hughes, Arthur, *Testing for Language Teachers. 2nd Ed* (London: Cambridge University Press, 2005).

whether true or false statement. The example of the true or false only have two possible answers they are True or false.

c) **Matching Test**

Matching test consists of matching questions that has to match on the pairing answers on the test that called stems. It includes of the easy to score and also quick to discuss. In this type the students have more time to do and respond the test, because some options will be similar with the other and they will get more mistakes if they get wrong of the one.

Test test that used in the research is the objective test, it meant that the test was answered by choosing the answers of the available answers. It was a multiple choice test, it consisted of a stem that is the question of the test, the choice of the right answer and the distractor that was the wrong answers with incomplete definitions.

2. **Criteria of a Good Test Item**

Test is used to measure the student's knowledge by assessing it during the process of learning. The test qualifies the materials that reflected on the test to get good score after getting the test. Test is called the good test if the test requires the criteria of a good test in the requirements statistically and non-statistically.

a. **Validity**

Validity is the important aspect in getting kinds of a good test. The validity of the test will be nothing if the test is not valid itself. The validity refers to the extent instrument of the objective to know the quality in the test suitable of the criteria.³⁴ To measure the test that the quality of

³⁴ Hatch, E & Farhady, H, *Research Design and statistics for applied linguistics*. Rowley, (Massachusetts: Newbury House, 1982), p.250.

validity is good or not, it should have terms of face validity, content validity, and construct validity.

1) Face Validity

Face validity concerns about the students and the teachers think of the test. It relates to the performance of the test how the test looks like.³⁵ Face validity assumes as the important comparison of the indications in the validity.

2) Content Validity

Content validity concerns about the correlation of the test and the materials in the test. It concerns about the sufficiently of the representative and the comprehensive about the test.³⁶ The content of validity reflects about the materials and knowledge that has been taught by the teacher during the learning in the classroom.

3) Construct Validity

Construct validity concerns about the specific objective in each aspect of the instruction of the test. The construct validity includes of the theory of the knowledge in the language.³⁷ It conducts of the test or final examination test that made by the teachers to measure the undersanding of students'. the test can be said as the construct validity if it conduct about the construct or theoretical ideas.

b. Reliability

Reliability refers to the consistency of measurement, to know the consistent score of the test or the evaluation

³⁵ Heaton, J. B, *Writing English Language Tests* (New York: Longman inc, 1990), P.159.

³⁶ Shohamy, EA *Practical Handbook in Language Testing for the Second Language Teacher*, Tel Aviv: Tel Aviv University, 1985, p.74.

³⁷ Ibid.

from the result of the measurement³⁸. The reliability will be reliable if the result of the test is consistently. It is the important one to know the result of the good test of the good validity.

c. Discriminating Power

Discrimination power refers to the test that tells about the items can discriminate the clever students and low students. Discrimination power tells of the extent quality of the options in the test to tell which items high or low students to answer the options.³⁹

d. Level of Difficulty

Level of difficulty refers to the percentage of the the correct items that gotten by the correct answers of the students'. The difficulty of the test will be seen from how easy or difficult items from the students' points of view.⁴⁰ Level of difficulty conducts of the difficulty and the easy items of test for the students. If the students answers the test correct mostly, so it means the test items range to easy items.

The criteria of a good test the researcher used were the validity, reliability, discrimination power, and level of difficulty. The researcher chose them as the point of the research because as the evaluating to know the quality of the questions that had made. It could be seen of the validity of the test that measure the extent of the validity of the questions, reliability of the test that is about the consistency of the research study, discrimination power that is about the item analysis, and the level of difficulty to know the level grade that used for its level. A good test it has fulfilled several

³⁸Linn, R., & Gronlund, *Measurement and evaluation in teaching 8th ed.* (New York: Macmillan, 2000). p 193.

³⁹ Shohamy, EA *Practical Handbook in Language Testing for the Second Language Teacher*, (Tel Aviv: Tel Aviv University, 1985),P.81.

⁴⁰ Ibid,79.

requirements of criteria of a good test statistically and non-statistically.

3. Iteman Software Program

a. Definition of Iteman Software Program

ITEMAN (Item and Test Analysis) is a Windows application designed to provide detailed reports of test items and analysis of their tests using classical test theory. ITEMAN is computer software especially made to analyze items or a test that is performed. ITEMAN is an empirical item analysis with a classical approach model that is useful for determining the quality of items or a test. ITEMAN (item and test analyzer) is a computer program used to analyze items classically⁴¹.

The purpose is to help the teachers to evaluate the quality of test items by examining their characteristics. This program includes a package of programs in the Micro Computer Adaptive Test (MicroCAT) developed by the Assessment Systems Corporation starting in 1982 and underwent revision in 1984, 1986, 1988, 1993, and 2013; starting from version 2.00 to version 4.30. Either Iteman 4.4 version in 2017.

Compared to version 3.5, ITEMAN version 4.4 is slightly different operation, but simpler, and even the reporting is very complete. ITEMAN version 4.4 produces text reports in RTF format, complete with graphs, narration, and tables.

Advantages of ITEMAN 4.4:

1. The result of the ITEMAN 4.4 can calculate as graphics. It generates the quantile for each items. The user see the result and control easily from the graphics.

⁴¹ Guyer, R and Thompson, N.A *User's Manual for Iteman: Assessment System Corporation*, (St Paul MN , 2011)

2. ITEMAN 4.4 can handle pretest or the trial that does not include in the final score statistically.
3. It has including the alpha result or the reliability a whole item.
4. The file that got prepared as a formal report and in comma separated. It makes the reader easy to read and evaluate in the file word and the report file is CSV format.
5. The scales of the score and sub score added in the output.
6. The score divides to two groups report file or result as the number.
7. The maximum number of score can be analyzed increase 10,000 items.
8. Multiple file created in the text editor or interactively within ITEMAN.⁴²

The results of the item analysis include the level of difficulty, differentiation, and distribution statistics of the answers. ITEMAN definition In addition to producing test item statistics, this program also produces test statistics which include test reliability, measurement error or standard error and score distribution.

4. English Language Skill Test

There are several skills in English to detect good tests based on the skills. They guide on the skills in English to make the examiner easy to establish the test for the examinee and make the examinee can develop their competence in English. "The effect of testing on teaching and learning is known as backwash..."⁴³

⁴² Sutrisno et al., ect, *ITEMAN: Item Klasik dan Analisis Uji*, (Semarang: CEP-CCDA, 2018)

⁴³ Hughes, Arthur. *Teaching for Language Teachers*, (Austalia: Cambridge University Press, 1989).

a. Testing Writing

There are some parts that the teacher have to understand the position problem in testing writing, They are:

- 1) The teacher has to set the writing test based on the population of task that the teacher expects the learners to be able to perform.
- 2) The students should elicit the samples of writing.
- 3) The essential is the task of the students' writing can and will be scored by the teacher.⁴⁴

in testing writing notices the use of the test for example to describe, to explain, to compare and contrast or argue. The examiner must observe some important things in testing writing such as:

- 1) Operations: the expressions are used in testing writing as expression of thanks, expression of opinions, expressions of apology, expression of information. Then the eliciting information, eliciting the directions, eliciting the service.
- 2) Text Type: The forms are used to see the purpose of writing as letter that includes the announcement, post that includes descriptive text, card that includes narrative text, or note that includes comment.
- 3) Addressees to test writing : The topic of writing such as medical, healthy, social interaction, education or many more.

b. Testing Oral Ability

The Problem in testing oral ability (speaking) is the same as testing writing, The different is only at the examinee performs the skill in spoken.⁴⁵ The task should elicit behavior which can be scored by the examiner.

⁴⁴ Ibid,75.

⁴⁵ Ibid,101.

This one is the important specification ‘Oral Interaction’ test at intermediate level serve as the illustration:

- 1) Operation : The operation to take the oral interaction is as the following function *the first* is expressing which includes thanks, opinion, comment, attitude, confirmation, apology, want/need, complains, or reason. *The second* is Narrating which tells the sequence of events. *The third* is Eliciting which the concerning of the talk as the information, direction, service, clarification, help, permission and all areas above. *The fourth* is directing which the purpose of oral interaction as ordering, instructing (how to), persuading, advising, warning. *The last* is reporting as description, comment, or decision.
- 2) Text Types : kinds of test in spokes as dialog or monologue.
- 3) Addresses and Topics: in oral test is not too specific not like in writing test.
- 4) Criteria Level of Performance in Speaking: in spoken test grammar is not specified as the content, there is no reference of vocabulary and pronunciation.
- 5) It does not meant there are no requirements of oral test just to be dealt with separately as part of statement of criteria level as :
 - a) The accuracy of the pronunciation must clear to list.
 - b) Appropriacy of the use of language generally appropriate to functions.
 - c) Range of language is available to the candidates, it means the speaker is able to express himself not stuck on searching words and too long briefing.

- d) Flexibility of the conversation adapt to new topic.
- e) Size of contribution in spoken may develop the utterances and the oral spoken.

The elicitation techniques that can be used to oral spoken are: ask and requests for pictures, information, interpreting, role play, comprehension, or discussion

c. Testing Reading

Testing reading is harder than testing speaking, when the reader reads some words we cannot observe but when the speaker speaks we can hear. Some important specification in kind of testing reading they are:⁴⁶

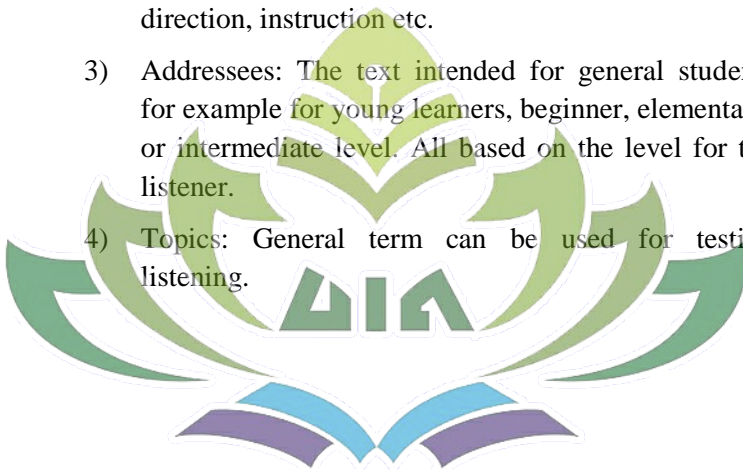
- 1) Operation: that the teacher have to notice to test reading skill, they can scan the text to locate the specific information, skimming, identifying stages the usage, identifying the examples that presented to support the argument and many other.
- 2) Type of Text: References of book/text that can be used is from textbook, novel, journal, letter, timetable, poem or future specified as newspaper report, newspaper advertisement, newspaper editorial etc. See the level of text that want to teach.
- 3) Addressees: Related to the text type see the quality text for the students in different culture and level.
- 4) Topics: Many topics that can discussed in English reading , as long as the text can be read and understand useful to reader and meaningful that means the book is able to use.

⁴⁶ Ibid,116.

d. Testing Listening

Many kinds of test to test listening, from many source of speaker start from radio, the speaker (teacher, lecturer or partner to talk), here are the special problem in constructing listening test, specify the students to do:⁴⁷

- 1) Operation: related to listening indicate as listening for the specific information, following directions, obtaining gist, following instructions or interpreting, recognize of the function of the structure.
- 2) Types of Text: The specified of test as monologue, dialogue, multi participants and further about direction, instruction etc.
- 3) Addressees: The text intended for general students for example for young learners, beginner, elementary, or intermediate level. All based on the level for the listener.
- 4) Topics: General term can be used for testing listening.



⁴⁷ Ibid,134.

REFERENCES

- Arikunto, s, 2006, *Prosedur Penelitian Suatu Pendekatan Praktik Edisi Revisi VI*, Jakarta: Rineka Cipta.
- Ary, Donald, 1990, *Introduction to Research in Education 4th*, New York: Harcourt Brace Collage Publishers.
- Brown, H. Douglas, 2004, *Language Assessment: Principle and Classroom Practices*, United States of America: Longman.
- Bachman, L. F, 1990, *Fundamental Considerations in language testing*, England: Oxford University Press.
- Colnelly, Lynne M, 2016, Trustworthiness in Qualitative Research, *Medsurd Nursing, Vol.25 No. 6*.
- Denzin, N, 1970, *Sociological Methods: A Source Book*, New Burnswick: Transaction Publishers.
- Dwi Rusmiana, Fatma, 2015, "The Test Analysis of 1st Semester Final Test of the Accounting Theory for Vocational Education: Case Study of SMK YPKK 1 Sleman Academic Year Of 2014/2015" . (The Research of Degree Bachelor of Education in Faculty of Economics Yogyakarta State University, Yogyakarta).
- Fiharsono, A, 2005, *An Analysis of the Use of Multiple Choice Items of Reading Test in UAN for Students of SMP*, Yogyakarta: Sanata Dharma University.
- Fulcher, Glenn, 2010, *Practical Language Testing*, Great Britain: Hachette UK Company.
- Gronlund, N. E, 1967, *Measurement and Evaluation in Teaching*, New York: Mac. Milan Co.
- Guyer, R and Thompson, N.A, 2011, *User's Manual for Iteman: Assessment System Corporation*, Minnesota: St Paul MN.
- Hatch, E & Farhady, H, 1982, *Research Design and statistics for Applied Linguistics*. Rowley, Massachusetts: Newbury House.
- Heaton, J. B, 1990, *Writing English Language Tests*, New York: Longman inc.

- Honey, P and Mumford, A, 1996, *The Manual of Learning Style*, 3rd edn, Maidenhead: Honey Publication.
- Hughes, Arthur, 1989, *Testing for Language Teachers*, New York: Cambridge University Press.
- Hughes, Arthur, 2005, *Testing for Language Teachers*. 2nd Ed, London: Cambridge University Press.
- Linn, R., & Gronlund, N, 2000, *Measurement and evaluation in Teaching 8th ed*, New York: Macmillan.
- Madsen S. Harold, 1983, *Techniques in Testing*, New York: Oxford University Press.
- Mark J. Gierl, Okan Bulut, Qi Guo, and Xinxin Zhang, 2017, "Developing, Analyzing, and Using Distractors for Multiple Choice Test in Education: Comprehensive". *Journal Education Research*, Vol. 87(December)
- Nitko, Anthony J, 1983, *Educational Tests and Measurement an Introduction*, New York: Harcourt Brace Jovanovich.
- Nofiyanti, 2011, "An Analysis on the Content Validity of The Summative Test for the First Year Students of Junior High School". (The Research of the Degree in English Language Education UIN Syarif Hidayatullah Jakarta, Jakarta).
- Nuttal, C, 1985, *Teaching Reading Skills in Foreign Language*, London: Heineman Education Book.
- O'Dweyer Aidar, 2012, "A Teaching Practices Review Of The Use Of Multiple Choice Questions for Formative and Summative Assessment of Student Work on Advanced Undergraduate and Postgraduate Modules in Engineering". *Journal AISHE*, Vol 4 No 1, Spring.
- Osterlind, S , 1998, *Constructing Test Items and Multiple Choice, Constructed Response, Performance and other Formats*, New York: Kluwer Academic Publishers.
- Polit, Desine F, 2020, *Essentials of Nursing Research*, Gold Coast: Wolters Kluwer Health.
- Shohamy, E, 1985, *A Practical Handbook in Language Testing for the Second Language Teacher*, Tel Aviv: Tel Aviv University.

Suparman, Ujang, 2011, "The Implementation of Iteman to Improve the Quality of English Test Items as a Foreign Language (An Assesment Analysis)". *Aksara Jurnal Bahasa*, Vol. XII. No. 1.

Suparman, Ujang, 2016, Analyzing the Quality of English Test Items of Daily, Mid Semester and Final School Examinations in Bandar Lampung, A paper presented in International Conference on Educational Evaluation and Research (ICERE), May 29 and 30.

Sutrisno et al, 2018, *ITEMAN: Item Klasik dan Analisis Uji*, Semarang : CEP-CCDA.

Arifin, Zaenal, 2010, *Metodologi Penelitian Pendidikan Filosofi, Teori dan Aplikasinya*, Surabaya: Lentera Cendekia.

